

## **Supporting Material**

### **Hemagglutinin receptor specificity and structural analyses of aerosol transmissible H5N1 viruses**

Robert P de Vries<sup>1</sup>, Xueyong Zhu<sup>2</sup>, Ryan McBride<sup>1</sup>, Alan Rigter<sup>3</sup>, Anthony Hanson<sup>4</sup>, Gongxun Zhong<sup>4</sup>, Masato Hatta<sup>4</sup>, Rui Xu<sup>2</sup>, Wenli Yu<sup>2</sup>, Yoshihiro Kawaoka<sup>4-7</sup>, Cornelis AM de Haan<sup>3</sup>, Ian A Wilson<sup>2,8</sup>, James C Paulson<sup>1,2\*</sup>

<sup>1</sup>Departments of Cell and Molecular Biology, and Chemical Physiology and <sup>2</sup> Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA.

<sup>3</sup>Virology Division, Department of Infectious Diseases & Immunology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, the Netherlands.

<sup>4</sup>Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, 575 Science Drive, Madison, WI 53711, USA

<sup>5</sup>ERATO Infection-Induced Host Responses Project, Japan Science and Technology Agency, Saitama 332-0012, Japan

<sup>6</sup>Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan

<sup>7</sup>Department of Special Pathogens, International Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108- 8639, Japan

<sup>8</sup>Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA.

\*Corresponding author: James C. Paulson  
10550 N. Torrey Pines Road, MB209, La Jolla, CA 92037, USA.  
Tel. +18587849634; fax: +185857849690  
E-mail address: [jpaulson@scripps.edu](mailto:jpaulson@scripps.edu)

## Material and Methods

### *Structural analyses*

The purified VN1203 HA RBS 3-mutant (N158D, N224K and Q226L) and VN1203 transmissible HA 4-mutant (N158D, N224K, Q226L and T318I) in 10 mM Tris, pH 8.0, 50 mM NaCl were concentrated to 7.0 mg/ml and 11.0 mg/ml, respectively, and subjected to crystallization trials using the automated Rigaku CrystalMation system at the Joint Center for Structural Genomics. The crystals used for data collection were grown by the sitting drop vapor diffusion method at 22°C with a reservoir solution containing 0.1 M Hepes, pH 7.5, 0.2 M KF, 22% (w/v) PEG 3350 for VN1203 HA 3-mutant or 0.1 M Tris, pH 8.0, 0.2 M KF, 20% (w/v) PEG 3350 for VN1203 HA 4-mutant. The crystals were flash-cooled in liquid nitrogen using 20% ethylene glycol (v/v) in mother liquor as cryoprotectant.

Structures were determined by molecular replacement with the program Phaser (1) using the VN1203 HA structure (PDB codes 3GBM) as a model. Initial rigid body refinement and restrained refinement were performed with program REFMAC5 (2) and PHENIX (3). Additional positive electron density was observed near all 5 potential *N*-glycosylation sites in the various HA monomers. The stem region of both mutant HAs showed higher flexibility in the crystal as indicated by higher *B*-values compared to the wild type, which had an antibody bound to the stem region of the HA (Fab CR6261).

**Supplementary Table 1.** Glycans imprinted on the microarray.

Glycan #	Name	Structure
1	Gal $\beta$ (1-4)GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
2	Gal $\beta$ (1-4)GlcNAc $\beta$ (1-2)Man $\alpha$ (1-3)[Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Man $\alpha$ (1-6)]-Man $\beta$ (1-4)-GlcNAc $\beta$ (1-4)-GlcNAc $\beta$ -Asn-NH <sub>2</sub>	
3	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-6-O-sulfo-GlcNAc $\beta$ -propyl-NH <sub>2</sub>	
4	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-[Fuc $\alpha$ (1-3)]-6-O-sulfo-GlcNAc $\beta$ -propyl-NH <sub>2</sub>	
5	NeuAc $\alpha$ (2-3)-6-O-sulfo-Gal $\beta$ (1-4)-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
6	NeuAc $\alpha$ (2-3)-6-O-sulfo-Gal $\beta$ (1-4)-[Fuc $\alpha$ (1-3)]-GlcNAc $\beta$ -propyl-NH <sub>2</sub>	
7	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-3)-6-O-sulfo-GlcNAc $\beta$ -propyl-NH <sub>2</sub>	
8	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-Glc $\beta$ -ethyl-NH <sub>2</sub>	
9	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
10	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
11	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
12	NeuAc $\alpha$ (2-3)-GalNAc $\beta$ (1-4)-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
13	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-3)-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
14	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-3)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
15	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-3)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-3)-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
16	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-3)-GalNAc $\beta$ (1-3)-Gala $\alpha$ (1-4)-Gal $\beta$ (1-4)-Glc $\beta$ -ethyl-NH <sub>2</sub>	
17	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-3)-GalNAc $\alpha$ -Thr-NH <sub>2</sub>	
18	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-3)-[GlcNAc $\beta$ (1-6)]-GalNAc $\alpha$ -Thr-NH <sub>2</sub>	
19	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-6)-[Gal $\beta$ (1-3)]-GalNAc $\alpha$ -Thr-NH <sub>2</sub>	

Glycan #	Name	Structure
20	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-6)-[Gal $\beta$ (1-3)]-GalNAc $\alpha$ -Thr-NH <sub>2</sub>	
21	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-GalNAc $\alpha$ -Thr-NH <sub>2</sub>	
22	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-GalNAc $\alpha$ -Thr-NH <sub>2</sub>	
23	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-[NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-6)]-GalNAc $\alpha$ -Thr-NH <sub>2</sub>	
24	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-[NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-6)]-GalNAc $\alpha$ -Thr-NH <sub>2</sub>	
25	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Mana(1-3)-[NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Mana(1-6)]-Man $\beta$ (1-4)-GlcNAc $\beta$ (1-4)-GlcNAc $\beta$ -Asn-NH <sub>2</sub>	
26	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Mana(1-3)-[NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Mana(1-6)]-Man $\beta$ (1-4)-GlcNAc $\beta$ (1-4)-GlcNAc $\beta$ -Asn-NH <sub>2</sub>	
27	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Mana(1-3)-[NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Mana(1-6)]-Man $\beta$ (1-4)-GlcNAc $\beta$ (1-4)-GlcNAc $\beta$ -Asn-NH <sub>2</sub>	
28	NeuAc $\alpha$ (2-3)-[GalNAc $\beta$ (1-4)]-Gal $\beta$ (1-4)-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
29	NeuAc $\alpha$ (2-3)-[GalNAc $\beta$ (1-4)]-Gal $\beta$ (1-4)-Glc $\beta$ -ethyl-NH <sub>2</sub>	
30	Gal $\beta$ (1-3)-GalNAc $\beta$ (1-4)-[NeuAc $\alpha$ (2-3)]-Gal $\beta$ (1-4)-Glc $\beta$ -ethyl-NH <sub>2</sub>	
31	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-[Fuc $\alpha$ (1-3)]-GlcNAc $\beta$ -propyl-NH <sub>2</sub>	
32	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-3)-[Fuc $\alpha$ (1-4)]-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-[Fuc $\alpha$ (1-3)]-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
33	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-[Fuc $\alpha$ (1-3)]-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-[Fuc $\alpha$ (1-3)]-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
34	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-[Fuc $\alpha$ (1-3)]-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-[Fuc $\alpha$ (1-3)]-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-[Fuc $\alpha$ (1-3)]-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	

Glycan #	Name	Structure
35	NeuGca(2-3)-Galβ(1-4)-GlcNAcβ-ethyl-NH <sub>2</sub>	
36	NeuAca(2-6)-Galβ(1-4)-6-O-sulfo-GlcNAcβ-propyl-NH <sub>2</sub>	
37	NeuAca(2-6)-Galβ(1-4)-Glcβ-ethyl-NH <sub>2</sub>	
38	NeuAca(2-6)-Galβ(1-4)-GlcNAcβ-ethyl-NH <sub>2</sub>	
39	NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ-ethyl-NH <sub>2</sub>	
40	NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ-ethyl-NH <sub>2</sub>	
41	NeuAca(2-6)-GalNAcβ(1-4)-GlcNAcβ-ethyl-NH <sub>2</sub>	
42	NeuAca(2-6)-[Galβ(1-3)]-GalNAcα-Thr-NH <sub>2</sub>	
43	NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-6)-[Galβ(1-3)]-GalNAcα-Thr-NH <sub>2</sub>	
44	NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ(1-6)-[Galβ(1-3)]-GalNAcα-Thr-NH <sub>2</sub>	
45	NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-3)-GalNAcα-Thr-NH <sub>2</sub>	
46	NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ(1-3)-GalNAcα-Thr-NH <sub>2</sub>	
47	NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-3)-[NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-6)]-GalNAcα-Thr-NH <sub>2</sub>	
48	NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ(1-3)-[NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ(1-6)]-GalNAcα-Thr-NH <sub>2</sub>	
49	Galβ(1-4)-GlcNAcβ(1-2)-Mana(1-3)-[NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-2)-Mana(1-6)]-Manβ(1-4)-GlcNAcβ(1-4)-GlcNAcβ-Asn-NH <sub>2</sub>	
50	NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-2)-Mana(1-3)-[Galβ(1-4)-GlcNAcβ(1-2)-Mana(1-6)]-Manβ(1-4)-GlcNAcβ(1-4)-GlcNAcβ-Asn-NH <sub>2</sub>	
51	GlcNAcβ(1-2)-Mana(1-3)-[NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-2)-Mana(1-6)]-Manβ(1-4)-GlcNAcβ(1-4)-GlcNAcβ-Asn-NH <sub>2</sub>	
52	NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-2)-Mana(1-3)-[NeuAca(2-6)-Galβ(1-4)-GlcNAcβ(1-2)-Mana(1-6)]-Manβ(1-4)-GlcNAcβ(1-4)-GlcNAcβ-Asn-NH <sub>2</sub>	

Glycan #	Name	Structure
53	NeuAc $\alpha$ (2-6)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Man $\alpha$ (1-3)-[NeuAc $\alpha$ (2-6)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Man $\alpha$ (1-6)]-Man $\beta$ (1-4)-GlcNAc $\beta$ (1-4)-GlcNAc $\beta$ -Asn-NH <sub>2</sub>	
54	NeuAc $\alpha$ (2-6)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Man $\alpha$ (1-3)-[NeuAc $\alpha$ (2-6)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Man $\alpha$ (1-6)]-Man $\beta$ (1-4)-GlcNAc $\beta$ (1-4)-GlcNAc $\beta$ -Asn-NH <sub>2</sub>	
55	NeuGc $\alpha$ (2-6)-Gal $\beta$ (1-4)-GlcNAc $\beta$ -ethyl-NH <sub>2</sub>	
56	NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Man $\alpha$ (1-3)-[NeuAc $\alpha$ (2-6)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Man $\alpha$ (1-6)]-Man $\beta$ (1-4)-GlcNAc $\beta$ (1-4)-GlcNAc $\beta$ -Asn-NH <sub>2</sub>	
57	NeuAc $\alpha$ (2-6)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Man $\alpha$ (1-3)-[NeuAc $\alpha$ (2-3)-Gal $\beta$ (1-4)-GlcNAc $\beta$ (1-2)-Man $\alpha$ (1-6)]-Man $\beta$ (1-4)-GlcNAc $\beta$ (1-4)-GlcNAc $\beta$ -Asn-NH <sub>2</sub>	

**Supplementary Table 2.** Data collection and refinement statistics for VN1203 HA mutants.

Data set	VN1203 HA 3-mutant	VN1203 HA 4-mutant
Space group	P1	P1
Unit cell (Å)	<i>a</i> = 117.9, <i>b</i> = 118.1, <i>c</i> = 273.8	<i>a</i> = 117.2, <i>b</i> = 118.9, <i>c</i> = 273.4
Unit cell (deg.)	$\alpha$ = 91.5 $\beta$ = 90.2 $\gamma$ = 119.9	$\alpha$ = 88.4 $\beta$ = 89.7 $\gamma$ = 60.3
Resolution (Å) <sup>a</sup>	50.0-3.20 (3.27-3.20)	50.0-2.95 (3.02-2.95)
X-ray source	APS 23ID-B	APS 23ID-B
Unique reflections	199,766	236,232
Redundancy <sup>a</sup>	1.9 (1.6)	1.9 (1.7)
Average <i>I</i> / $\sigma$ ( <i>I</i> ) <sup>a</sup>	7.1 (1.1)	8.5 (1.1)
Completeness (%) <sup>a</sup>	93.6 (83.4)	89.7 (77.4)
<i>R</i> <sub>sym</sub> <sup>a,b</sup>	0.13 (0.68)	0.13 (0.71)
HA monomers in a.u.	15	15
<i>V</i> <sub>m</sub> (Å <sup>3</sup> /Da)	3.9	3.9
Reflections in refinement	199,509	235,905
Refined atoms	61,068	61,083
Refined waters	0	0
<i>R</i> <sub>cryst</sub> <sup>c</sup>	0.213	0.235
<i>R</i> <sub>free</sub> <sup>d</sup>	0.261	0.281
<i>B</i> -values (Å <sup>2</sup> )		
Protein	97	91
Wilson <i>B</i> -values (Å <sup>2</sup> )	63	64
Ramachandran plot (%) <sup>e</sup>	92.9, 0.1	93.8, 0.1
r.m.s.d. bond (Å)	0.012	0.011
r.m.s.d. angle (deg.)	1.56	1.46
PDB code	4N5Y	4N5Z

<sup>a</sup> Parenthesis denote outer-shell statistics.

<sup>b</sup>  $R_{\text{sym}} = \sum_{hkl} \sum_i |I_{hkl,i} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_i I_{hkl,i}$ , where  $I_{hkl,i}$  is the scaled intensity of the  $i^{\text{th}}$  measurement of reflection  $h, k, l$ ,  $\langle I_{hkl} \rangle$  is the average intensity for that reflection.

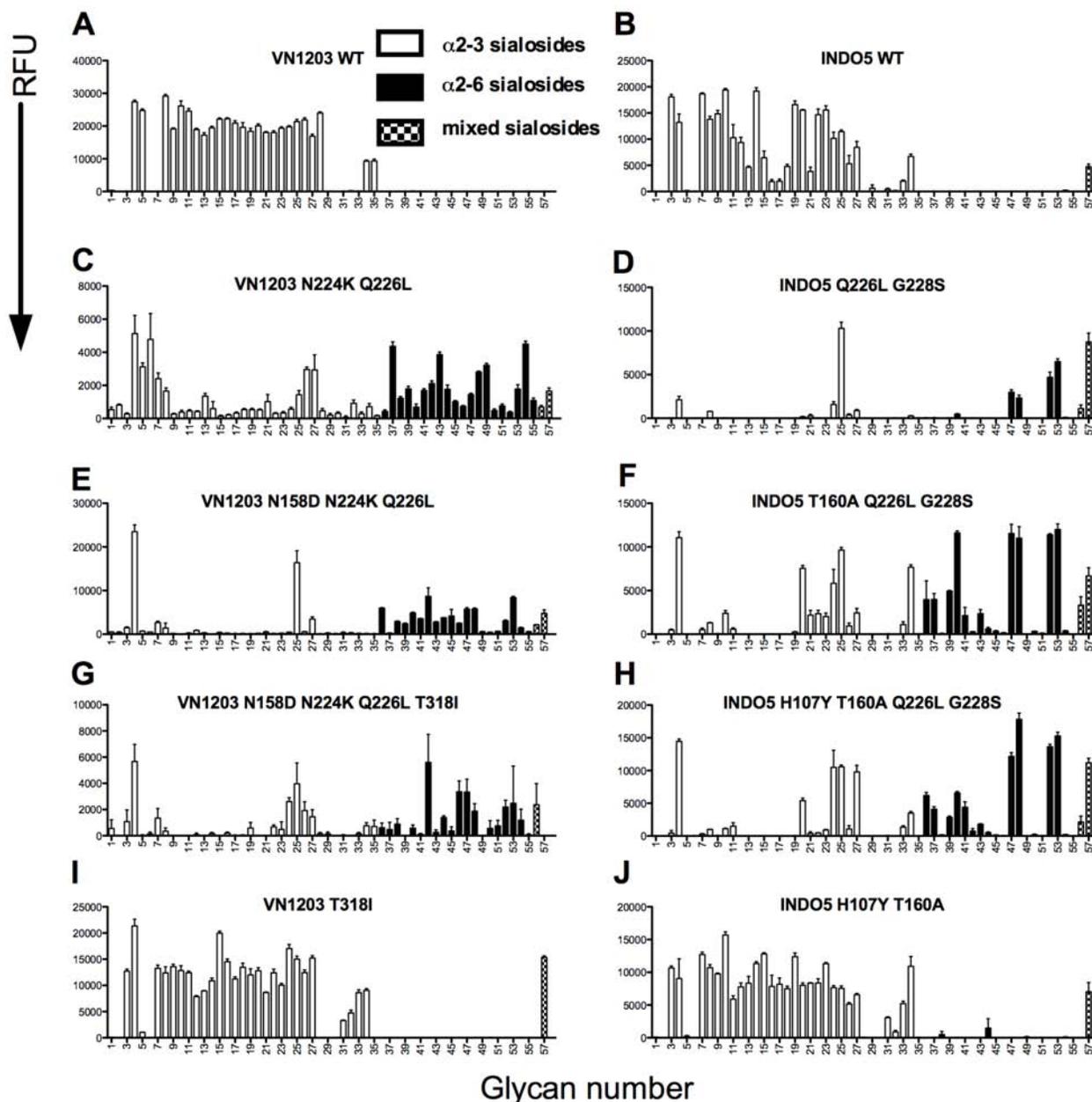
<sup>c</sup>  $R_{\text{cryst}} = \sum_{hkl} |F_o - F_c| / \sum_{hkl} |F_o|$ , where  $F_o$  and  $F_c$  are the observed and calculated structure factors.

<sup>d</sup>  $R_{\text{free}}$  was calculated as for  $R_{\text{cryst}}$ , but on 5% of data excluded before refinement.

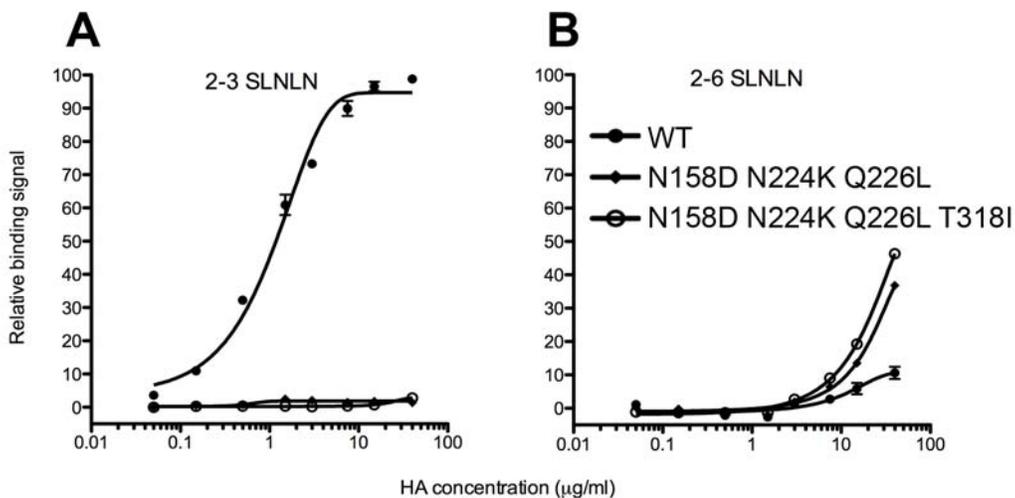
<sup>e</sup> The values are percentage of residues in the favored and outliers regions analyzed by MolProbity (4).

**Supplementary Figure 1. Full array data of receptor specificity of aerosol transmissible recombinant VN1203 and INDO5 H5 hemagglutinins.**

The receptors specificity of recombinant H5 HAs representing a series of WT and receptor mutants of VN1203 (*left panels*) and INDO5 (*right panels*), was assessed on glycan microarrays described in Figure 1. The  $\alpha$ -2,3 sialosides are in white bars and the  $\alpha$ -2,6 sialosides are in black bars. Shown are results for the VN1203 WT (A) and INDO5 WT (B) and VN1203 HAs with the RBS mutations (C and D). HA with the RBS mutants and removal of the glycosylation site (E and F), and addition of the stalk mutations, T318I (G) and H107Y (H), that were also introduced into the HA without the RBS mutations (I and J). The glycans are shown in Table S1.



**Supplementary Figure 2.** ELISA-based plate assay of VN1203 HA WT, RBS and stalk mutations from a baculovirus expression system. Mutants and WT to 3'-SLNLN (structure **10**) (A) and 6'-SLNLN (structure **39**) (B).



## References

1. **McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. 2005.** Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 61:458-464.
2. **Murshudov GN, Vagin AA, Dodson EJ. 1997.** Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240-255.
3. **Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010.** PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213-221.
4. **Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010.** MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66:12-21.